# Data-Driven Modeling: Two Methodological Generalizations

*Time-Varying* Inputs and *Time-Resolved* Outputs


***CHRISTOS GEORGAKIS***

Systems Research Institute for Chemical and Biological Processes

& Department of Chemical and Biological Engineering

**TUFTS UNIVERSITY**

**MEDFORD MA 02155, USA**

# What I am Going to Tell You?



- Generalization of the Design of Experiments
  - Design of Dynamic Experiments
  - Dynamic Response Surface Models
- Use them to:
  - Model Process Not Well Understood
    - Mostly Batch but also … Continuous Processes
  - Optimize them
    - Almost as Well as with a Knowledge-Driven Model (KDM)
  - Even … Proceed towards a KNM
- Industrial Applications

# The Generalization of DoE



- DoE a Very Powerful Methodology 50 Years Young!
  - Full and Fractional Factorial Designs, ANOVA
  - RSM: Interpolative and Linear and Nonlinear Models
    - Linear in Parameters
- Two Major Limitations of DoE
  - Inputs Do NOT Vary with Time
    - Why Keep Reaction Temperature Constant?
    - Why Keep Co-reactant Flow Constant?
  - Outputs Measurements at End of Experiment
    - We Take On-Line Spectral and Other Measurement VERY frequently.
- ***Our Answer is DoDE and DRSM***

# The DoDE Approach

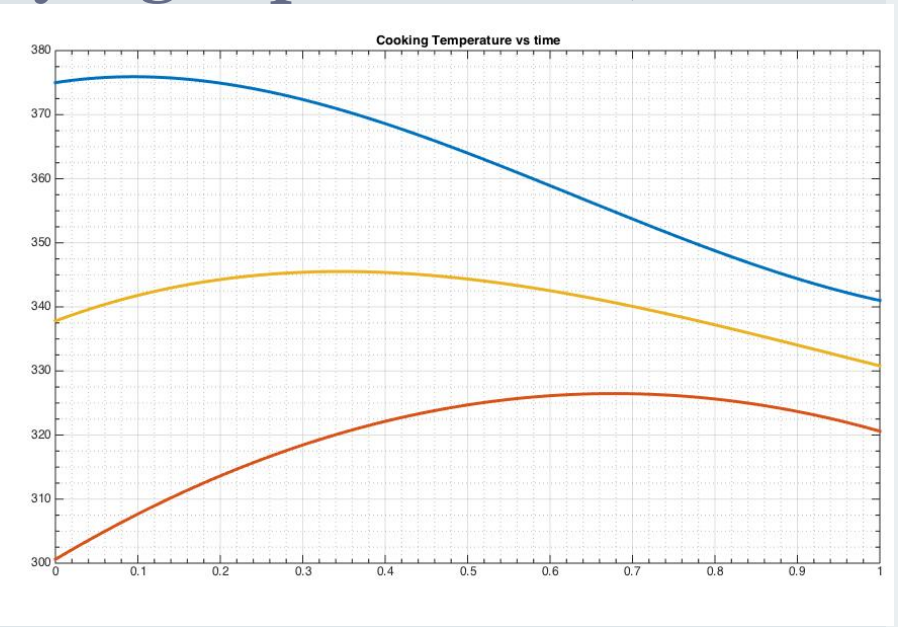


- Applicable to ANY Time-Varying Input Factor, u(t)
  - Define Coded variable, z(τ)

$$u(\tau) \triangleq u_0(\tau) + \Delta u(\tau) z(\tau)$$

$$z(\tau) = \frac{u(\tau) - u_0(\tau)}{\Delta u(\tau)}, \quad \begin{cases} u_0(\tau) = \dfrac{u_{\max}(\tau) + u_{\min}(\tau)}{2} \\ \Delta u(\tau) = \dfrac{u_{\max}(\tau) - u_{\min}(\tau)}{2} \end{cases}$$

$$-1 \le z(\tau) \le +1, \quad \tau = t / t_b$$




- Parameterize Input: z(τ)
  - Using: $P_i(\tau)$=Shifted Legendre Polynomials

$$z(\tau) = \sum_{i=1}^{n} x_i P_{i-1}(\tau).$$

$$P_0(\tau) = 1,\ P_1(\tau) = -1 + 2\tau,\ P_2(\tau) = 1 - 6\tau + 6\tau^2,\ \ldots$$

Orthogonality: $\int_0^1 P_i(\tau) P_j(\tau) d\tau = 0$ for $i \ne j$

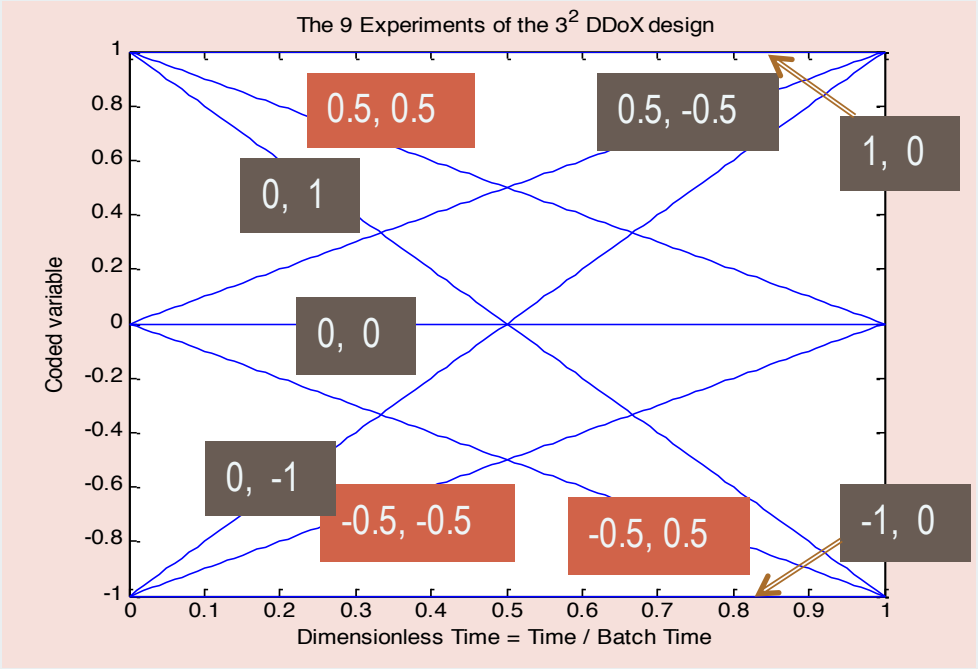
- Dynamic Sub-factors: $x_1, x_2, \ldots, x_n;\ -1 \le x_1 \pm x_2 \pm \ldots \pm x_n \le +1$

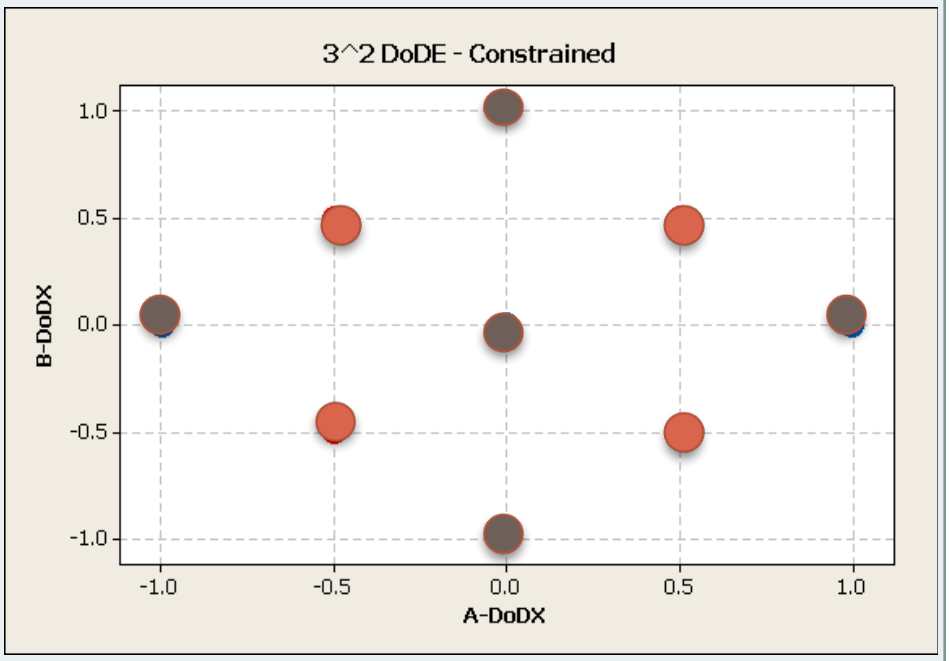
# DoDE with n=2: a $3^2$ Design



- Dynamic Factor: $z(\tau)$
  - Dynamic Subfactors: $x_1$ and $x_2$

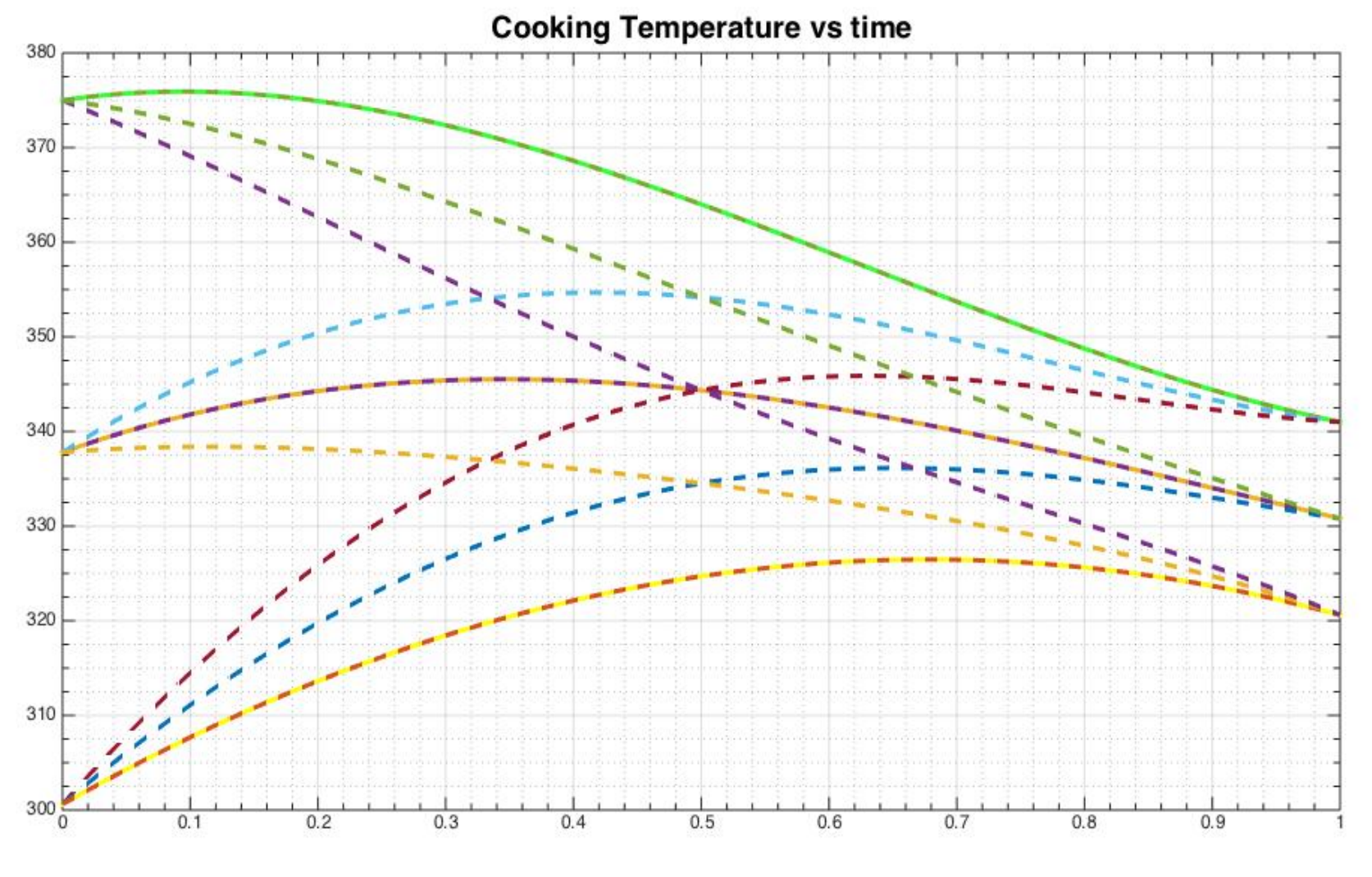
$$z(\tau) = x_1 P_0(\tau) + x_2 P_1(\tau) = x_1 + x_2(2\tau - 1); \quad \& \quad -1 \le x_1 \pm x_2 \le +1$$



# The nine (9) runs within the Region



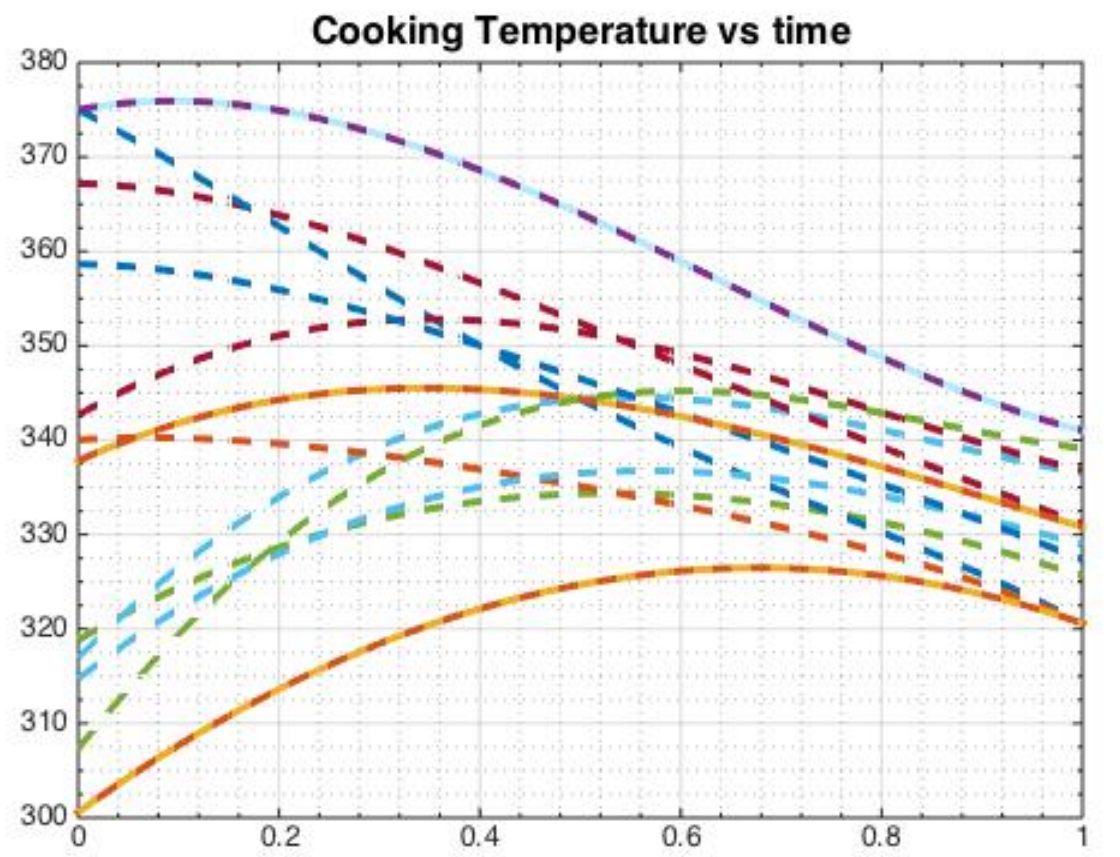


# Quadratic Time Profiles



- **The $2^3=8$ *Full Factorial DoDE***
- Dynamic Factor: $z(\tau)$
  - Dynamic Subfactors: $x_1$, $x_2$ and $x_3$

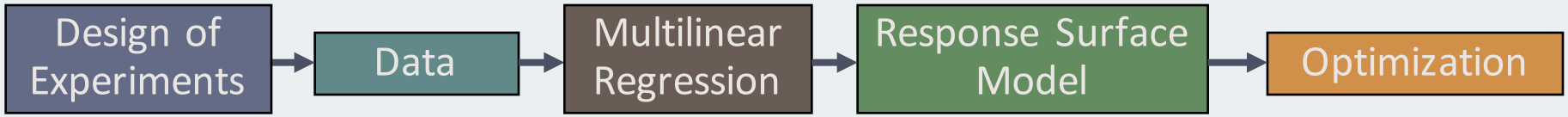
$$z(\tau) =$$
$$= x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau) =$$
$$= x_1 + x_2(2\tau - 1) + x_3(1 - 6\tau + 6\tau^2)$$
$$\&$$
$$-1 \le x_1 \pm x_2 \pm x_3 \le +1$$
$$\text{so that}$$
$$-1 \le z(\tau) \le +1$$



# DoE & DoDE - Response Surface Models



## ❖ The DoE Steps




## ❖ Response Surface Model (RSM)

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j>i}^{n} \beta_{ij} x_i x_j + \sum_{\iota=1}^{n} \beta_{ii} x_i^2$$

## ❖ Design of ***Dynamic*** Experiments: **The Same!**

Parameterize Time-Varying Input $z(\tau)$; $(\tau = t/t_b)$

$$z(\tau) = \sum_{i=0}^{n} x_i P_i(\tau); P_i(\tau) = \text{Shifted Legandre Polynomials}$$

**PARAMETERIZE, DoDE experiments, RSM, OPTIMIZE**



Georgakis, C., (2009) ADCHEM Proceedings, Istanbul, July 2009.
Georgakis, C. (2013). Ind. Eng. Chem. Res. **52** (35): 12369-12382.

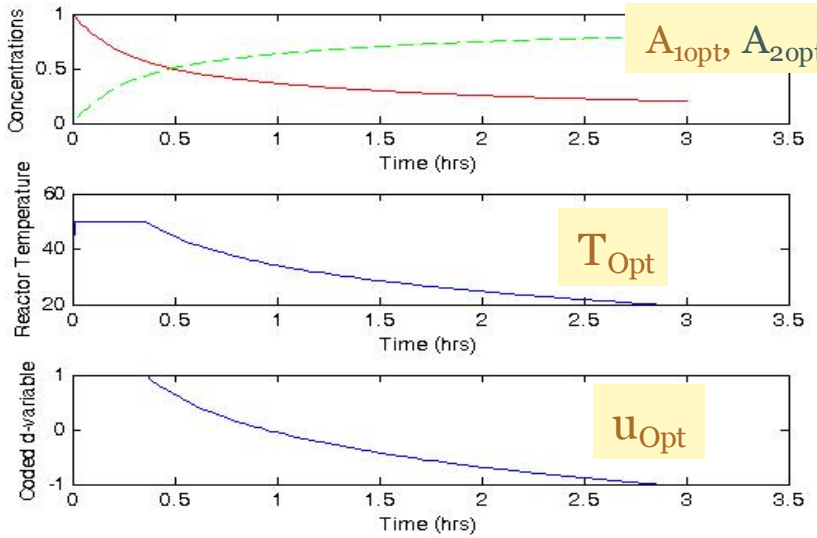
# DoDE Example: Batch Reactor



- Batch Reversible Reaction [15 < T < 50 °C]
  - $A_1 \rightleftharpoons A_2$ $k_i = k_{io}\exp(-E_i/RT)$ with $E_2 > E_1$

***Model-based Optimum:*** Decreasing Temperature Profile
***Optimum Conversion=74.57%*** at $t_b$=2.0 hr

DoE
or
DoDE
???



# Reactor Optimization via DoE & DoDE



- Single Factor: Reactor Temperature
  - Data: Conversion at 2hr + Error (±3%)
  - Five **DoE** Experiments at T=15, 32.5 (3), and 50 °C
    - **T constant with time!**
  - Nine **DoDE** Experiment (T(t) linear in Time)
    - Between 15 and 50°C
- Optimization: Maximum Conversion
  - DoE Optimum: x=71.44 at T*= 36.25 °C
  - DoDE Optimum: **x=74.32,** T* from 50 to 28°C
  - Model-Based (True) Optimum = ***74.57%***

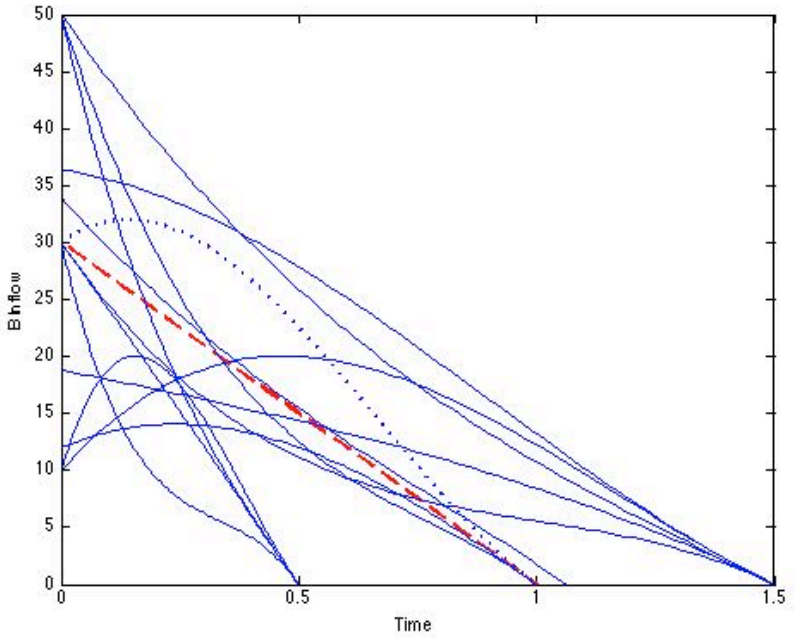
# DoDE on Isothermal Semi-Batch Reactor



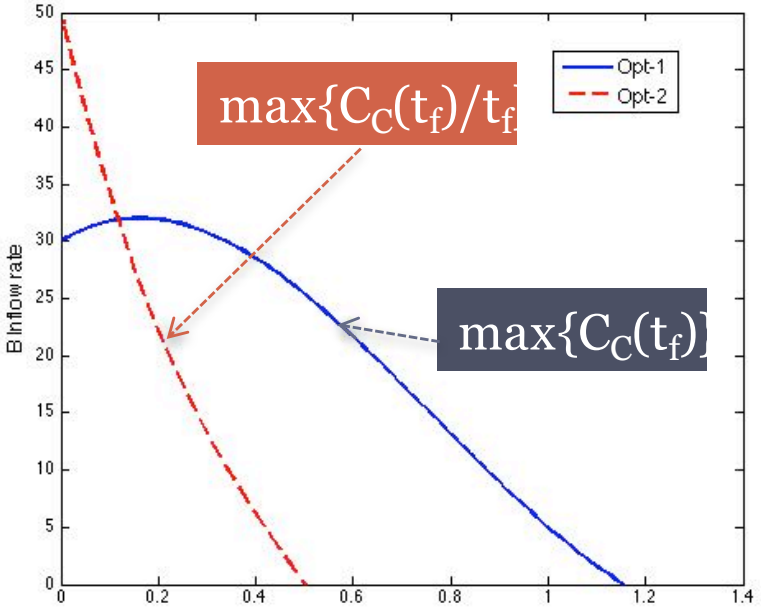
- Reaction Example:

$$Rxn1:\ A+B\rightarrow C,\ r_1=k_1C_AC_B,\ k_1=2l\,mol^{-1}h^{-1}$$
$$Rxn2:\ 2B\rightarrow D,\quad r_2=k_2C_B^2,\quad k_2=1l\,mol^{-1}h^{-1}$$
$$Rxn2:\ C\rightarrow E,\quad r_2=k_3C_C,\quad k_3=1h^{-1}$$

DoDE Runs: Feeding B




Optimal Runs



# Sepracor Pharmaceutical Reaction System



## Asymmetric Catalytic Hydrogenation

$$\text{Reactant} \xrightarrow[\text{Hydrogenation}]{\text{Catalytic}} \text{trans-Product}$$

$$\text{Reactant} \xrightarrow[\text{Hydrogenation}]{\text{Catalytic}} \text{cis-Product}$$

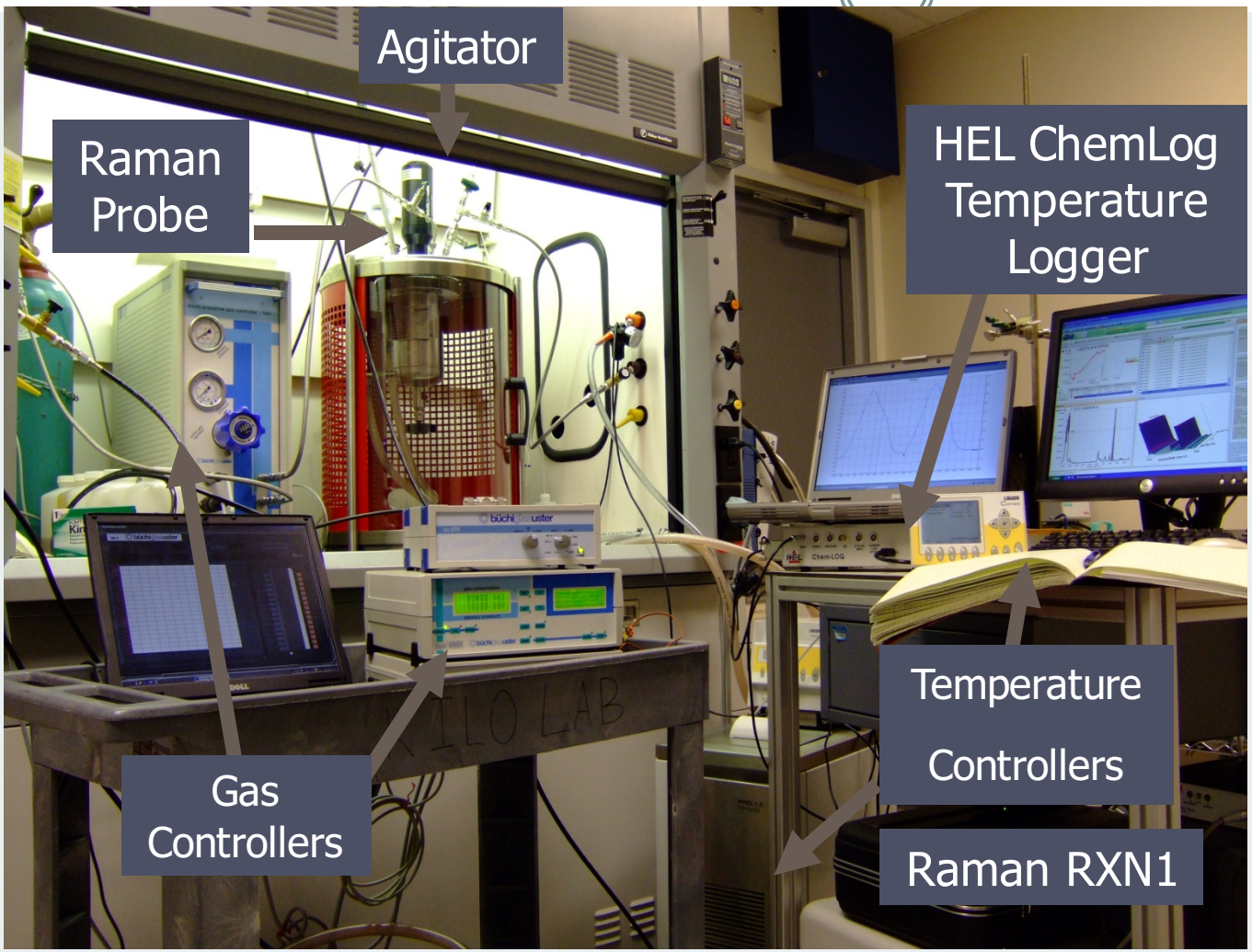
*Project Specific Goals:*

- Optimize Reaction Conditions
  - Selectivity of Asymmetric Hydrogenation
  - Minimize Catalyst Loading
- Performance Criterion
  - Profit = Value of Product- Cost of Reactants

Experiments and Analysis performed by Fenia Makrydaki, PhD candidate

# Sepracor Experimental System






## Advantages

- ✓ Accurate Measurements
- ✓ Precise Pressure Control
- ✓ H2 Consumption Monitoring
- ✓ Minimize Mass Transfer Limitations

# Design of Dynamic Experiments – DoDE

## $8=2^3$ experiments with 2 Levels & 3 (2+1dynamic) Factors Full Factorial



- ***Time Variant Experiments: Temperature Profile***
- ***Advantages: Additional degrees of freedom***

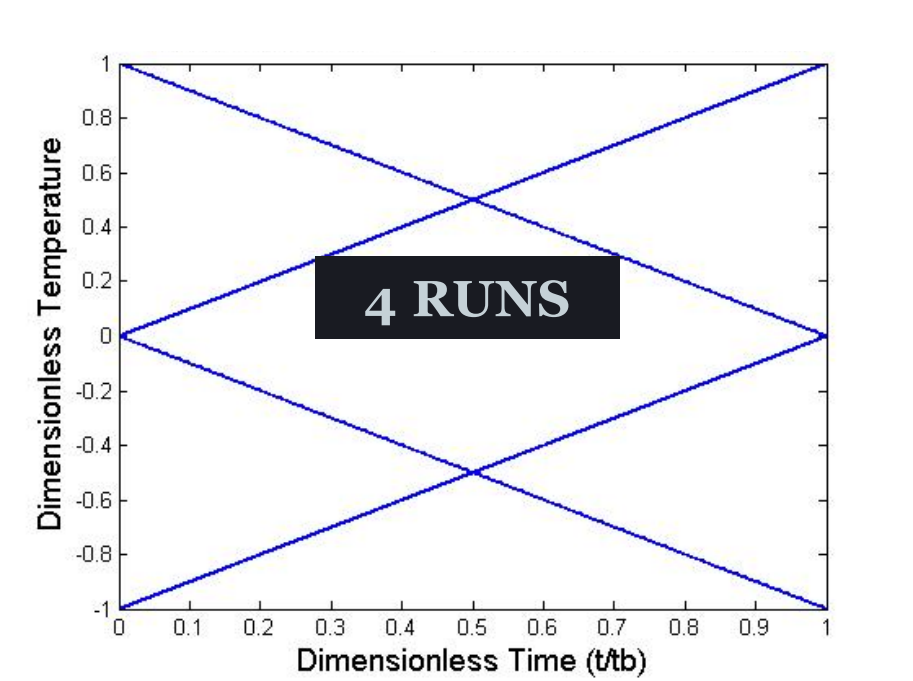



Figure A: 2 level, 2 factor, full factorial case

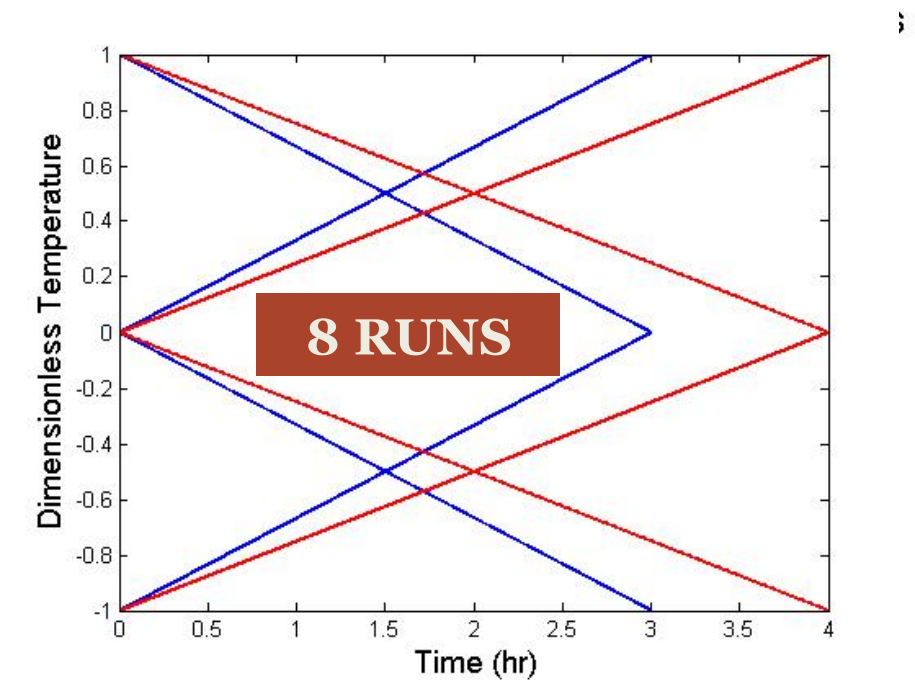



Figure B: 2 level, 2 factor, full factorial case for two time horizons.

Makrydaki, F., Georgakis, C., & Saranteas, K.: DYCOPS 2010, Leuven, Belgium, July 5-7, 2010

# DoE Design Table & Responses



- D-Optimal Experimental Design -17 Runs
  - with 3 Center Points

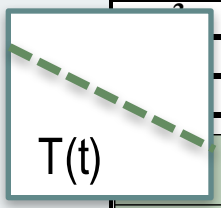
| Run | $x_1$ (T) | $x_2$ (RE) | $x_3$ (CL) | $x_4$ (BT) | DE (%) | Y (%) | PI ($/l) |
|---|---|---|---|---|---|---|---|
| **1** | -1 | 1.67 | 0 | -1 | 97.4 | 85.5 | 562.2 |
| **2** | 1 | 1.67 | -1 | 1 | 93.7 | 75.7 | 897.7 |
| **3** | -1 | -1 | -1 | -1 | 97.7 | 98.6 | 155.7 |
| **4** | -1 | 1.67 | 1 | -2.25 | 97.3 | 96.1 | 668 |
| **5** | 1 | 1.67 | -1 | -1 | 90 | 39.7 | -16.2 |
| **6** | -1 | -1 | -1 | -1 | 97.7 | 98.6 | 155.7 |
| **BEST DoE Run** | | | | | | | |
| **7** | **-1** | **1.67** | **-1** | **0** | **96.8** | **95.4** | **726.1** |
| **12** | 1 | 1.67 | 1 | 0 | 93.8 | 94.6 | 1242.9 |
| **13** | 0 | 1.67 | 1 | -1 | 95.7 | 85.7 | 714.1 |
| **14** | 1 | -1 | -1 | -2 | 94 | 74.7 | 174.7 |
| **15** | 0 | 0 | 0 | -2.33 | 94.9 | 97.4 | 497.8 |
| **16** | 0 | 0 | 0 | -2.25 | 95.8 | 97.7 | 501 |
| **17** | 0 | 0 | 0 | -2.33 | 95.9 | 97.5 | 499.3 |

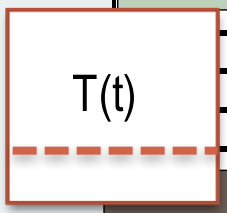
# *DoDE* Design Table & Responses

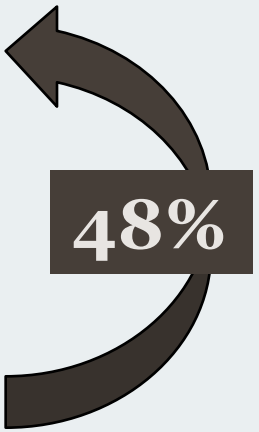


- 3 Static & 2 Dynamic Factors, 21 Runs +3 CPs

| Run | $x_1$ ($a_0$) | $x_2$ ($a_1$) | $x_3$ (RE) | $x_4$ (CL) | $x_5$ (BT) | DE (%) | Y (%) | PI ($/l) |
|---|---|---|---|---|---|---|---|---|
| 1 | -0.88 | 0.13 | -1 | 1 | -2.5 | 97.3 | 98.7 | 142.1 |
| 2 | 0.06 | 0.06 | -1 | 1 | -2.5 | 96.5 | 97.4 | 191.6 |
| | -0.88 | 0.13 | 1.67 | 1 | -2 | 97.8 | 97.2 | 684.6 |
| | 0.15 | 0.15 | -1 | -1 | -2.42 | 96.1 | 98 | 212.8 |
| | 0.94 | -0.06 | -1 | -1 | -2.5 | 93.6 | 96.8 | 288.3 |
| | 0.5 | -0.5 | 1.67 | -1 | -1 | 96.4 | 19.1 | -534.3 |
| **BEST DoDE Run** | | | | | | | | |
| **9** | **0** | **-1** | **1.67** | **-1** | **1** | **96.1** | **83** | **1076.2** |
| | 0.5 | -0.5 | 0 | 1 | -1.5 | 96.8 | 97 | 656.5 |
| | -0.5 | 0.5 | 1.67 | -1 | -1 | 96.7 | 73.4 | 436.9 |
| | 0.5 | -0.5 | 1.67 | 1 | 1 | 94.6 | 96.2 | 1294.9 |
| | -0.25 | -0.25 | -1 | -1 | -2 | 96.7 | 98.3 | 213.7 |
| | -0.72 | 0.28 | -1 | -1 | -0.75 | 97.6 | 98.8 | 155.8 |
| **BEST DOE Run** | | | | | | | | |
| **7** | **-1** | **--** | **1.67** | **-1** | **0** | **96.8** | **95.4** | **726.1** |
| **19** | -0.5 | -0.5 | 1.67 | 1 | 0 | 97.1 | 96.7 | 941.3 |
| **20** | -1 | 0 | -1 | 1 | -2.33 | 97.8 | 98.8 | 142.2 |
| **21** | -0.11 | -0.11 | 0 | 0 | -2.17 | 96.5 | 98.2 | 509.1 |
| **22** | 0 | 0 | 0 | 0 | -2.33 | 94.9 | 97.4 | 497.8 |
| **23** | 0 | 0 | 0 | 0 | -2.25 | 95.8 | 97.7 | 501 |
| **24** | 0 | 0 | 0 | 0 | -2.33 | 95.9 | 97.5 | 499.3 |



# DoDE for Penicillin Fermentation



- Use Dynamic Model to Simulate Experiments

$$x_1 = V: \quad \frac{dV}{dt} = u(t)$$

$$x_2 = x: \quad \frac{dx}{dt} = \mu x - \frac{x}{s_f V} \qquad \mu = \mu_{\max} \frac{s}{k_x x + s}$$

$$x_3 = s: \quad \frac{ds}{dt} = -\frac{\mu x}{Y_{x/s}} - \rho \frac{x}{Y_{p/s}} - \frac{m_s s}{k_m + s} x + (s_f - s)\frac{u}{V}$$

$$x_4 = p: \quad \frac{dp}{dt} = \rho x - kd - \frac{p}{V} u \qquad \rho = \rho \max \left( \frac{s}{k_p + s + s^2 / k_{in}} \right)$$

Riascos & Pinto, *Chem. Eng. Sci.* **99** (2004) 23-34
Bajpai & Reuss, *Biotechnol. Bioeng.* 23 (1981) 717

# DoDE Optimization Task



- Calculate U(t) to Maximize $V(t_f)p(t_f)$
  - With Constraint $V(t_f)$=10 lt; V(0)=7 lt
- Model-Based Optimum: 85.4 gr
  - (Riascos & Pinto, *Chem. Eng. Sci.* **99** (2004) p.23)
- Design of the DoDE Experiments
  - Batch Time: $100<t_b<160$ hr
  - Initial Biomass Concentration: 0.5-1.5 *gr/lt*
  - Feeding Profile: Quadratic
- Optimize Process Using the DoDE-RSM
  - **Best DoDE run= 85.07 gr**

Small Difference

# What Did I Tell You so Far?



- Design of Dynamic Experiments (DoDE)
  - First Generalization of DoE
    - New Set of Inputs: TIME-VARYING
  - Effective Optimization of Processes
  - SMALL distance from Model-Based Optimum

# **PART B:** Dynamic RSM (DRSM)



- Use Time-Resolved Output Data
- Classical RSM: $$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \beta_{ii} x_i^2$$
- Dynamic RSM: $$y(\tau) = \beta_0(\tau) + \sum_{i=1}^{n} \beta_i(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{ii}(\tau) x_i^2$$

K= # of C Measurements in Time

# of $\beta(\tau)$ functions=1+n+0.5n(n-1)+n

- Parameterization of $$\beta_{ij}(\tau) = \gamma_{ij,1} P_0(\tau) + \gamma_{ij,2} P_1(\tau) + \ldots + \gamma_{ij,R} P_{R-1}(\tau)$$

R= # of Polynomials

- Number of Model parameters: $$p = (1 + 2n + 0.5n(n-1))R$$
- Number of Data: $$d = (1 + 2n + 0.5n(n-1) + 6)K$$

R<K

# DRSM for Simple Batch Reaction



$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B$$

$$r = k_1[A] - k_2[B]$$

**Not enough measurements in time (K=3)**

$$k_1 = k_{10} \exp\left(-\frac{E_1}{RT}\right),$$

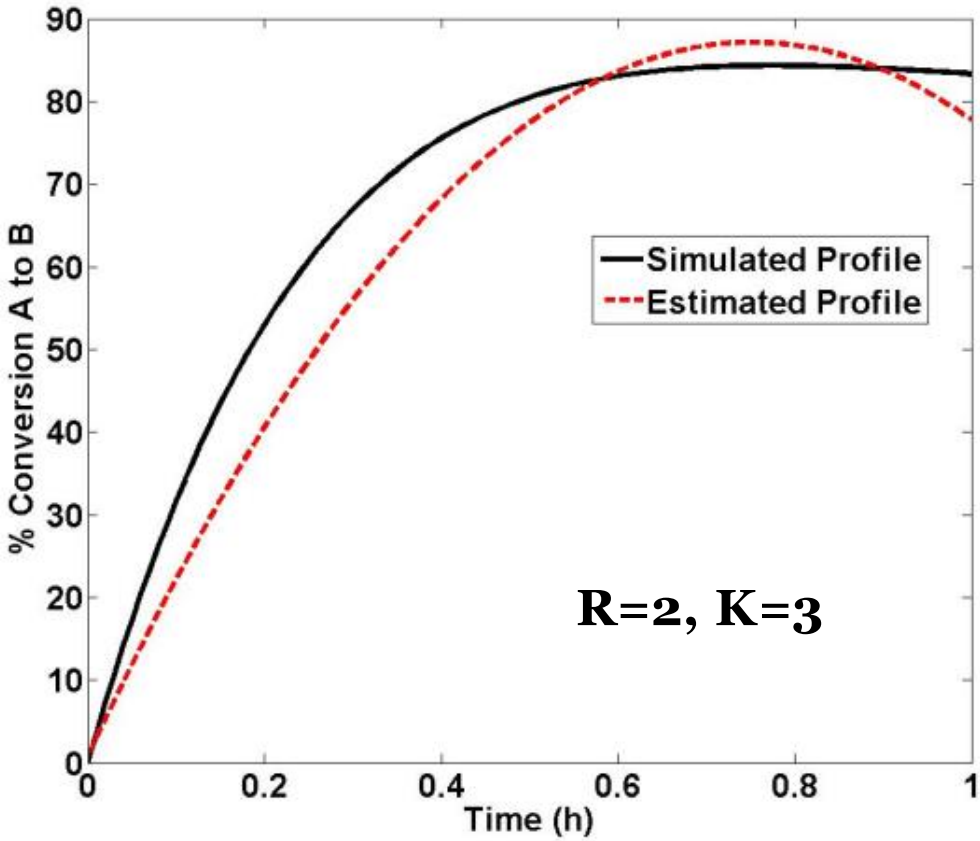
$$k_{10} = 1.32 \times 10^8 \text{ h}^{-1},$$

$$E_1 = 10{,}000 \text{ kcal},$$

$$k_2 = k_{20} \exp\left(-\frac{E_2}{RT}\right),$$

$$k_{20} = 5.25 \times 10^{13} \text{ h}^{-1},$$

$$E_2 = 20{,}000 \text{ kcal}$$



# Statistical Measure of Accuracy



- **Unmodeled Variance:**

$$\hat{SS}_{un}(R,K)=\frac{\sum_{i=1}^{M}\left[\sum_{k=1}^{K}\left\{y_{RSM,i}(\tau_k;R,K)-y_{\exp,i}(\tau_k)\right\}^2\right]}{\sum_{i=1}^{M}\left[\sum_{k=1}^{K}\left\{y_{\exp,i}(\tau_k)\right\}^2\right]}$$

- **Normal Variability:**

$$\hat{SS}_{err}=\frac{\sum_{i=1}^{n_{CP}}\left[\sum_{k=1}^{K}\left\{y_{0,i}(\tau_k)-\bar{y}_{0,i}(\tau_k)\right\}^2\right]}{\sum_{i=1}^{N_{CP}}\left[\sum_{k=1}^{K}\left\{\bar{y}_{0,i}(\tau_k)\right\}^2\right]}$$

- **Hypothesis Testing:**

$$\text{Null Hypothesis} ------ H_0: \quad \hat{SS}_{err}=\hat{SS}_{rgr}(R,K)$$

$$\text{Alternative Hypothesis} -- H_1: \quad \hat{SS}_{err}<\hat{SS}_{rgr}(R,K)$$

- ***F*-Statistic:**

$$F_0(R,K)=\frac{\hat{SS}_{un}(R,K)/n_1}{\hat{SS}_{err}/n_2}=\frac{\hat{SS}_{un}(R,K)/(MK-Q)}{\hat{SS}_{err}/K(n_{CP}-1)}$$

# Reactor Example: K=Measurements, R=Polynomials



**Table 3.** p-values of F-test for DRSM of A ⇄ B conversion in batch reactor

| R \ K | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 3 | | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 4 | | | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 5 | | | | **0.51** | 0.99 | **0.90** | **0.95** | **0.70** |
| 6 | | | | | **0.82** | **0.34** | **0.48** | **0.03** |
| 7 | | | | | | **0.47** | **0.47** | **0.01** |
| 8 | | | | | | | **0.58** | **0.01** |
| 9 | | | | | | | | **0.14** |

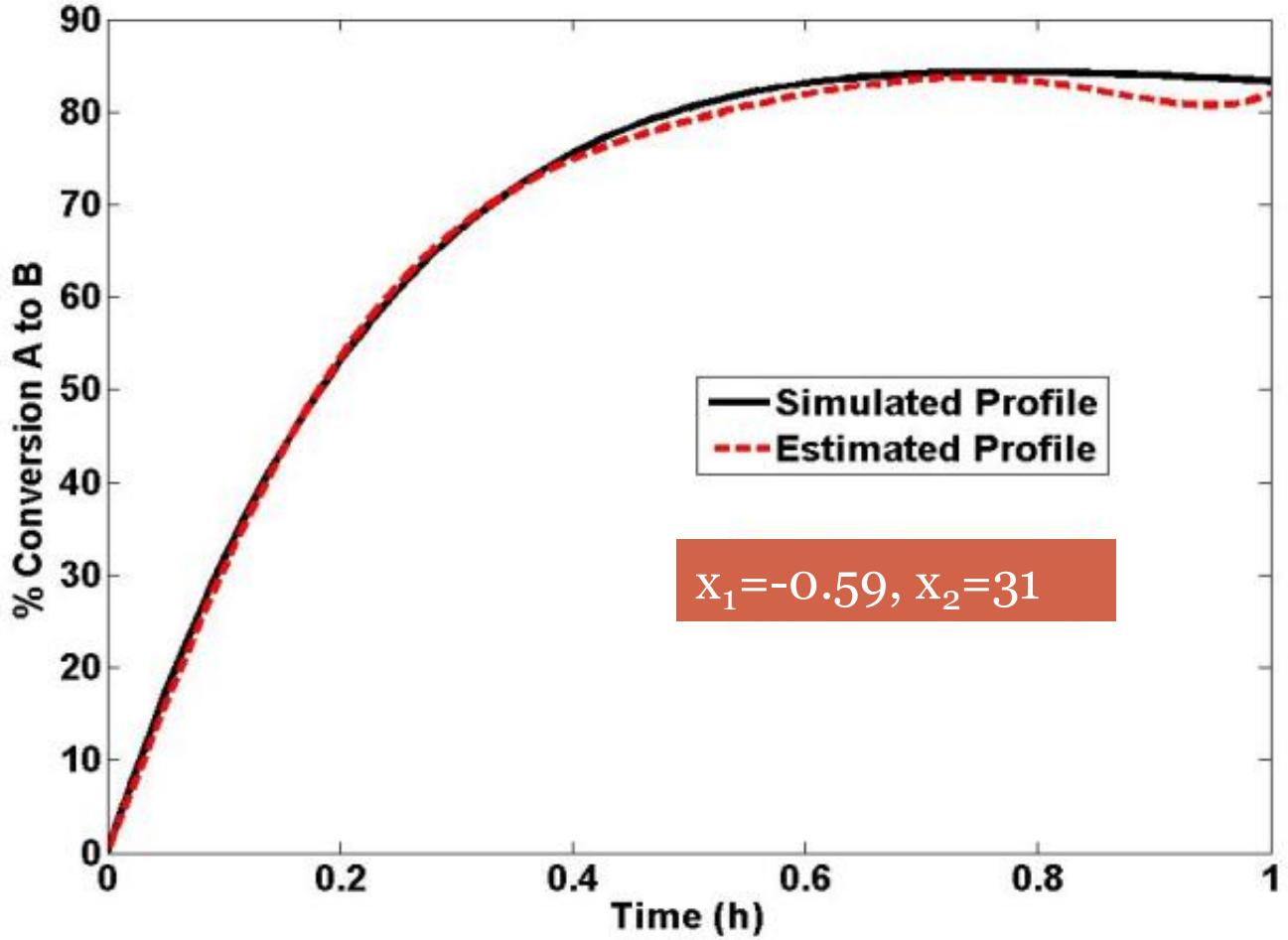
K = number of time-resolved measurements; R = number of polynomials

$if \quad p(R,K) \leq 0.95$ the Null Hypothesis **fails** to be rejected $\Rightarrow$ **Model GOOD**

$if \quad p(R,K) > 0.95$ the Null Hypothesis **is** rejected

# D-RSM Model (R=7, K=14)





# More Complex Semi-Batch Case



- Three inter-related reactions
- **C** is the desired product
  - **Reactant B is fed in semi-batch mode**

Rxn1: $A + B \rightarrow C,$ $r_1 = k_1[A][B]$ with $k_1 = 2$ 1 gmol $h^{-1}$

Rxn2: $2B \rightarrow D,$ $r_2 = k_2[B]^2$ with $k_2 = 1$ 1 gmol $h^{-1}$

Rxn3: $C \rightarrow E,$ $r_3 = k_3[C]$ with $k_3 = 1$ $h^{-1}$

- DRSMs for A(t), B(t), C(t), D(t), and E(t)

# Statistical Test of Goodness-of –Fit (GoF)



**Table 6.** Corresponding F-test p values for DRSM of product [C] in semi-batch 3-reaction network

| *R* \ *K* | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 3 | | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 4 | | | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 5 | | | | **2.33E-06** | **8.06E-06** | **2.89E-05** | **1.45E-02** | **2.22E-05** |
| 6 | | | | | **3.18E-26** | **1.22E-29** | **5.18E-27** | **6.14E-37** |
| 7 | | | | | | **6.89E-36** | **9.60E-38** | **1.07E-50** |
| 8 | | | | | | | **9.60E-38** | **1.07E-50** |
| 9 | | | | | | | | **1.07E-50** |

K = number of time-resolved measurements; R = number of polynomials

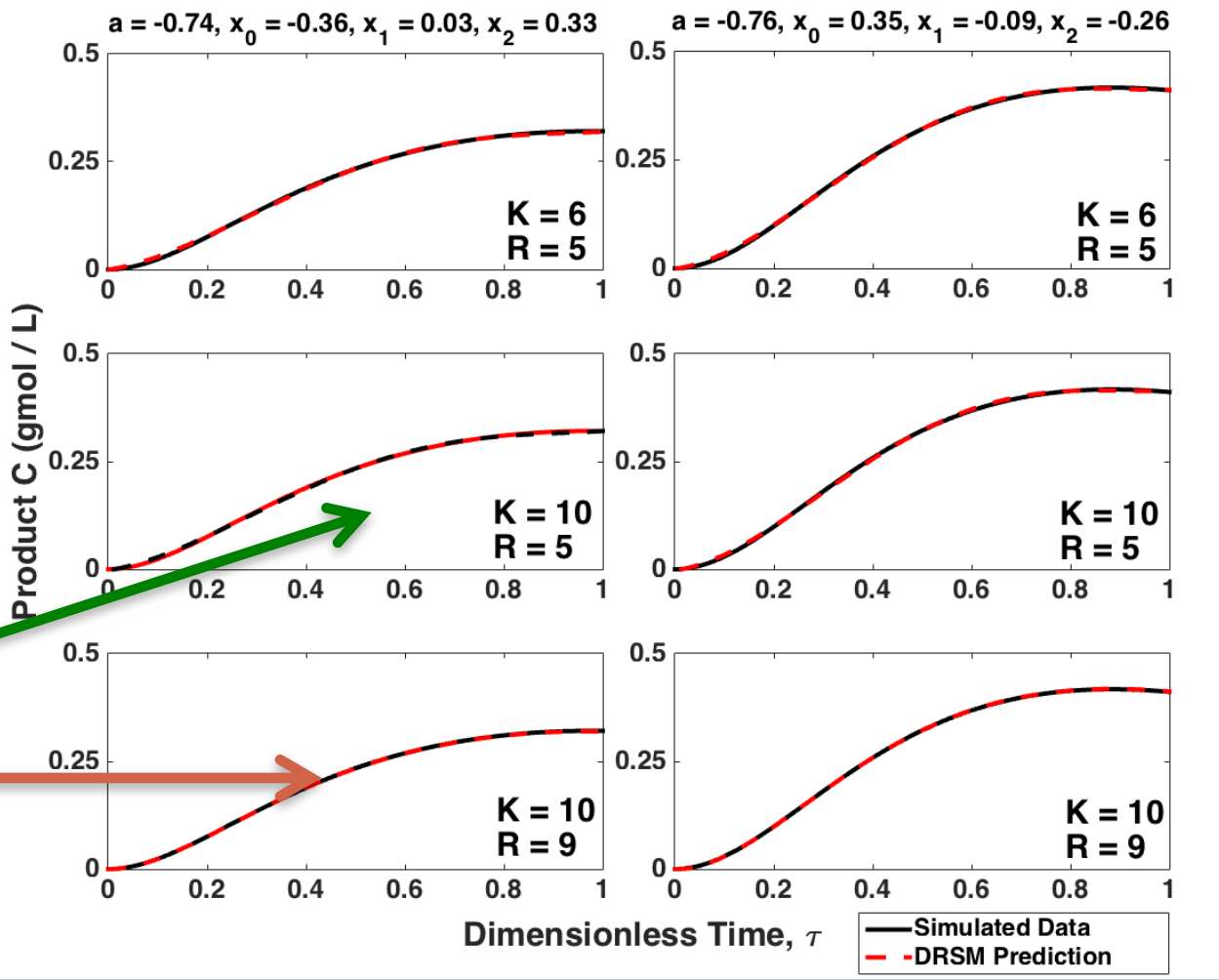
**Excellent Model**

# Some C(t) Profiles



- Excellent fits:

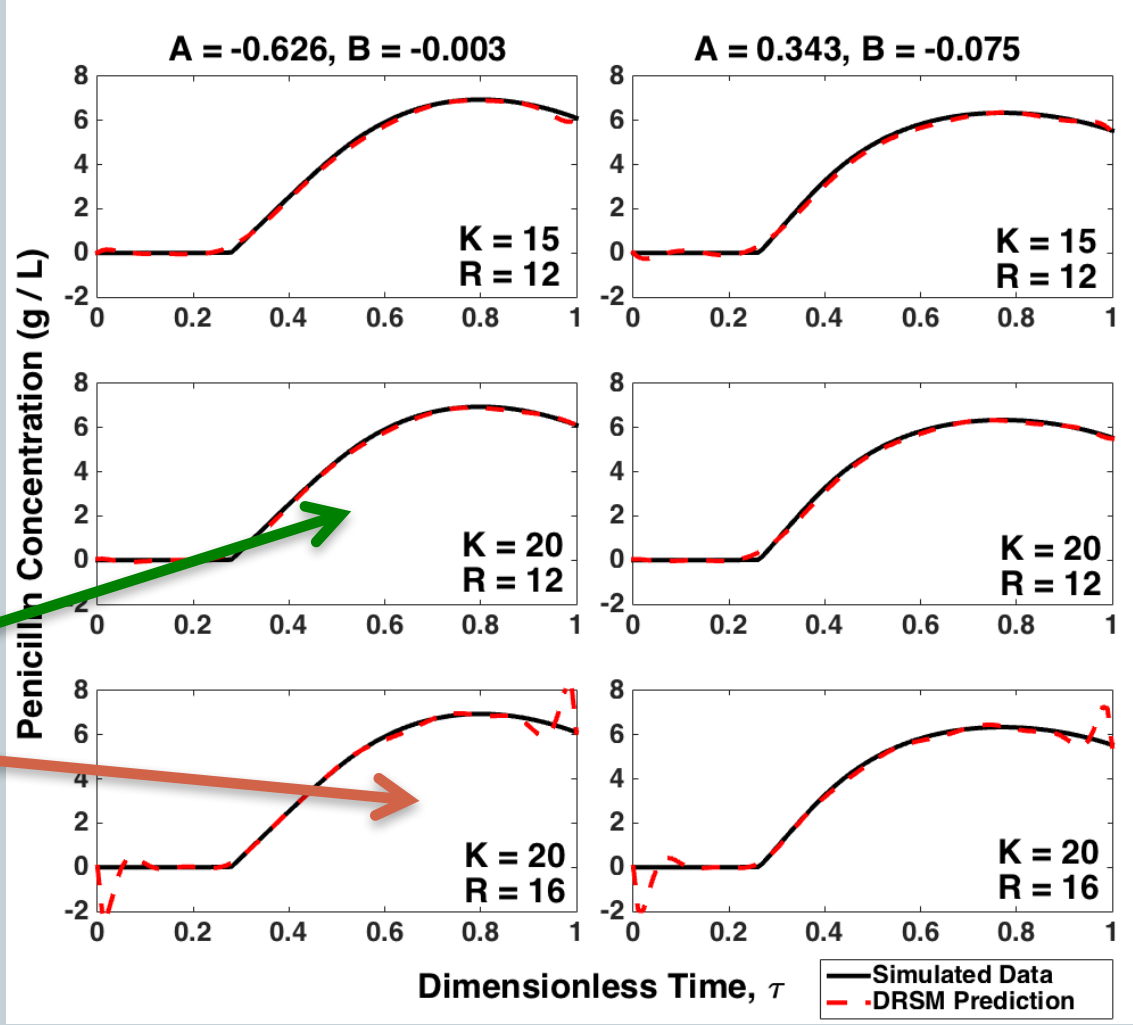
No Significant Difference Using Stepwise Regression



# Penicillin Concentrations



- Excellent fits,
  - Despite Challenge
- Watch Out for
  - Larger than Needed R
  - Compare
  - Despite SWR



# Part C: DRSM Usage → Door to Knowledge



- Revisit Semi-Batch Reactor Example
  - Five DRSMs at Hand

$$\Rightarrow \begin{cases} \mathrm{R}_1\colon\ A + B \rightarrow C \\ \mathrm{R}_2\colon\ 2B \rightarrow D \\ \mathrm{R}_3\colon\ C \rightarrow E \end{cases}$$

$$c_A(\tau) = \beta_{A0}(\tau) + \sum_{i=1}^{n} \beta_{Ai}(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{Aij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{Aii}(\tau) x_i^2$$

$$c_B(\tau) = \beta_{B0}(\tau) + \sum_{i=1}^{n} \beta_{Bi}(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{Bij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{Bii}(\tau) x_i^2$$

$$c_C(\tau) = \beta_{C0}(\tau) + \sum_{i=1}^{n} \beta_{Ci}(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{Cij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{Cii}(\tau) x_i^2$$

$$c_D(\tau) = \beta_{D0}(\tau) + \sum_{i=1}^{n} \beta_{Di}(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{Dij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{Dii}(\tau) x_i^2$$

$$c_E(\tau) = \beta_{E0}(\tau) + \sum_{i=1}^{n} \beta_{Ei}(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{Eij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{Eii}(\tau) x_i^2$$

- Can Calculate Derivatives wrt Time

$$\frac{dc_A(\tau)}{d\tau} = \frac{d\beta_{A0}(\tau)}{d\tau} + \sum_{i=1}^{n} \frac{d\beta_{Ai}(\tau)}{d\tau} x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{d\beta_{Aij}(\tau)}{d\tau} x_i x_j + \sum_{i=1}^{n} \frac{d\beta_{Aii}(\tau)}{d\tau} x_i^2$$

… for ALL experiments

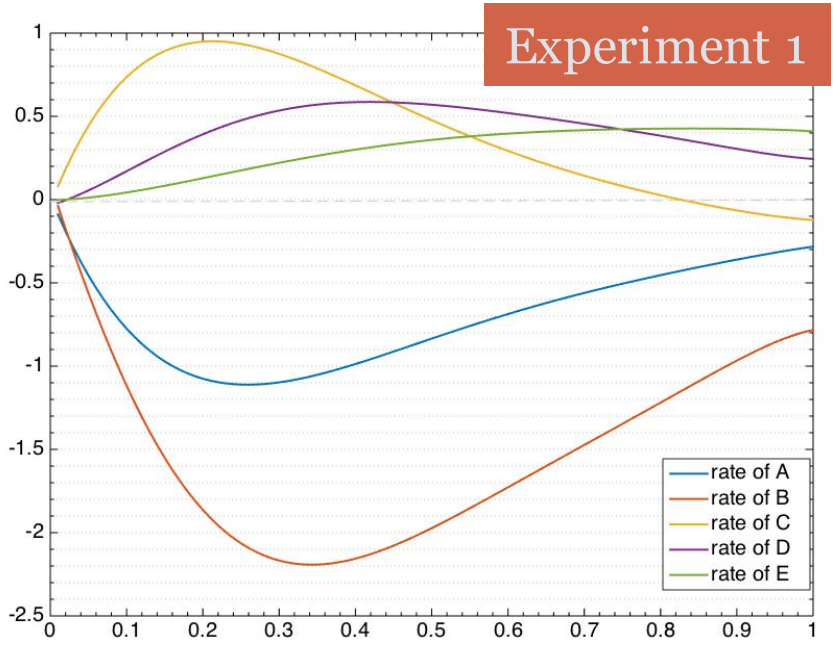
# Calculate Rate of Appearance (Disappearance)



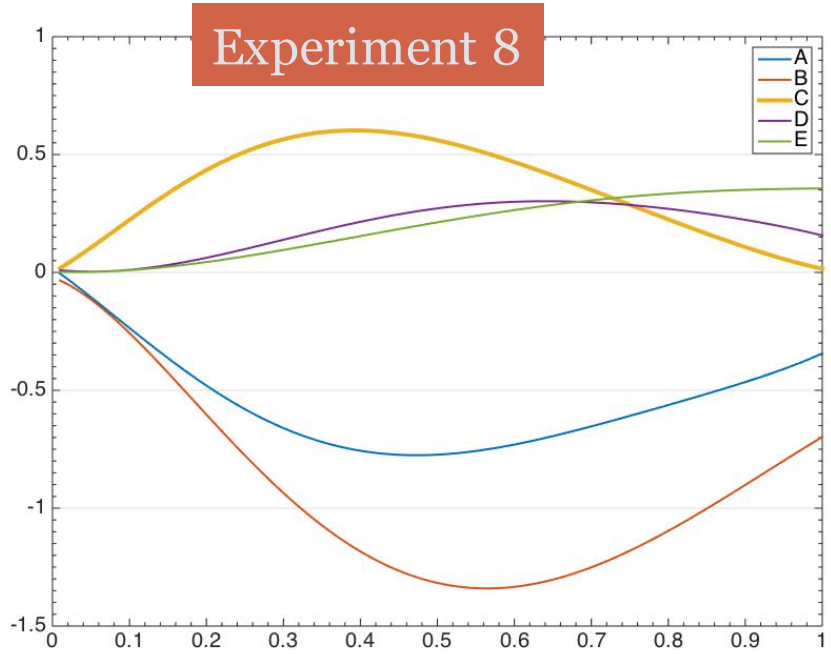
- Calculate at 100 time points in each Run:
  - τ=0.01, 0.02, …, 0.99, 1.00
- Can plot the Rates vs. Time
  - Can Understand what is Happening

$$r_A(\tau)=\frac{dc_A(\tau)}{d\tau}, r_C(\tau)=\frac{dc_C(\tau)}{d\tau},$$
$$r_D(\tau)=\frac{dc_D(\tau)}{d\tau}, r_F(\tau)=\frac{dc_E(\tau)}{d\tau}$$
$$\textit{but}\quad r_B(\tau)=\frac{dc_B(\tau)}{d\tau}-\frac{q_B(\tau)}{V}$$

Experiment 1




Experiment 8



# Discover the Stoichiometry



- Define Big Rate Data Matrix: DataM

$$\mathbf{DataM} = \begin{pmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_Q \end{pmatrix}, \quad \mathbf{D}_k = \text{Data from } k\text{-th Experiment}$$

$$Q = 1 + n + 0.5n(n-1) + n + 6(3)$$

$$\text{for } n = 2 \Rightarrow Q = 12(9)$$

$$\mathbf{D}_k = \begin{pmatrix} r_{kA}(0.01) & r_{kB}(0.01) & r_{kC}(0.01) & r_{kD}(0.01) & r_{kE}(0.01) \\ r_{kA}(0.02) & r_{kB}(0.02) & r_{kC}(0.02) & r_{kD}(0.02) & r_{kE}(0.02) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{kA}(1.00) & r_{kB}(1.00) & r_{kC}(1.00) & r_{kD}(1.00) & r_{kE}(1.00) \end{pmatrix}$$

- DataM is a 909 x 5 matrix !!!
- SVD of DataM

$$SVD(\mathbf{DataM}) = \mathbf{USV}^T$$

  - **S has three Dominant Singular Values** → three Reactions !!
  - **Matrix V is KEY to Stoichiometry**

# Testing Stoichiometries (measure A, B. C. D, and E)



- SVD Results

$$\mathbf{V}^T = \begin{pmatrix} 0.4072 & 0.8365 & -0.2557 & -0.2145 & -0.1516 \\ -0.2596 & 0.2086 & 0.7629 & -0.2333 & -0.5027 \\ 0.5955 & -0.2826 & 0.0140 & 0.4417 & -0.6084 \end{pmatrix}$$

$$\mathrm{R}_1\colon\ A + B \rightarrow C$$
$$\mathrm{R}_2\colon\ 2B \rightarrow D$$
$$\mathrm{R}_3\colon\ C \rightarrow E$$

- TEST: $\mathbf{N}_r = \mathbf{N}\mathbf{V}\mathbf{V}^T$; is $\mathbf{N}_r = \mathbf{N}$???
- Test TRUE Stoichiometry
  - Score: 99.74%

$$\mathbf{N} = \begin{pmatrix} -1 & -1 & 1 & 0 & 0 \\ 0 & -2 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

- Test *Incorrect* Stoichiometry
  - Score: 64.97%

$$\mathbf{N} = \begin{pmatrix} -1 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

- Can Test **One** Reaction at a Time

# Testing Stoichiometries (measure A, B. C. and D - **NO E**)



- SVD Results

$$\mathbf{V}^T = \begin{pmatrix} 0.4125 & 0.8456 & -0.2608 & -0.2164 \\ -0.2841 & 0.3233 & 0.8503 & -0.3029 \\ 0.7620 & -0.1172 & 0.4571 & 0.4435 \end{pmatrix}$$

$$R_1: A + B \rightarrow C$$
$$R_2: 2B \rightarrow D$$
$$R_3: C \rightarrow E$$

- Test TRUE Stoichiometry
  - Score: 99.67%

$$\mathbf{N} = \begin{pmatrix} -1 & -1 & 1 & 0 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

- Test *Incorrect* Stoichiometry
  - Score: 64.87%

$$\mathbf{N} = \begin{pmatrix} -1 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

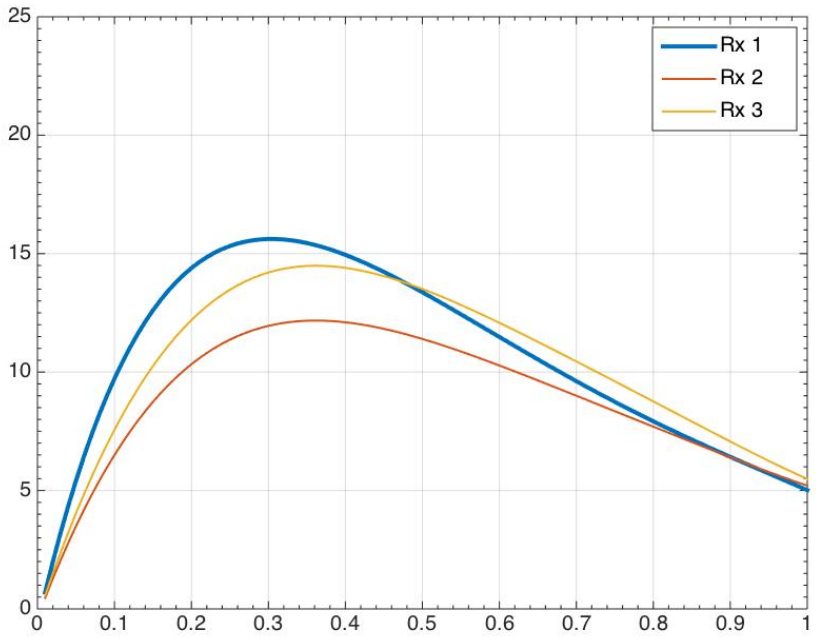
- Can Test **One** Reaction at a Time

# Calculate Reaction Rates

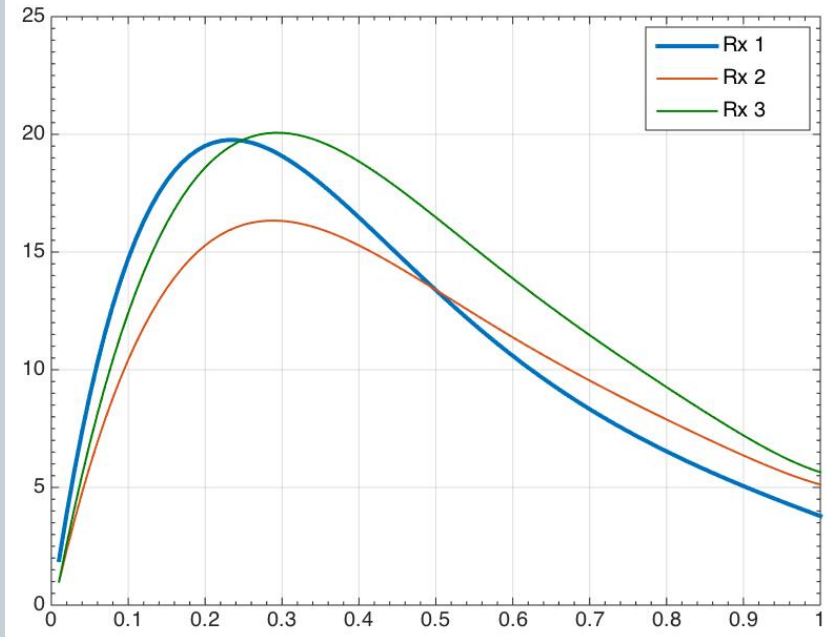


- From: $r_A(\tau), r_B(\tau), r_C(\tau), r_D(\tau), r_E(\tau)$
  - **TO:** $r_1(\tau), r_2(\tau), r_3(\tau)$

Reaction Rates Experiment 4




Reaction Rates Experiment 8



# Derive Kinetic Laws



- From: $r_1(\tau)$, $C_A(\tau)$, $C_B(\tau)$
  - TO: Kinetic Rates $r_1(\tau)=f(C_A(\tau), C_B(\tau))$
- Work in Progress
  - Stay … Tuned

# What did I just Tell you?



- DRSM is a Generalization of RSM
  - Using Time-Resolved Measurements
    - Excellent Approximation of Composition Profiles
  - Stepping Stone to Stoichiometry and a Kinetic Model

# Industrial Applications



- Dow Batch Polymerization Reactor
  - Use DoDE to Increase Productivity by 20%
    - Presented at the Houston AIChE Meeting
    - ***Will Give you the Highlights***
- Pfizer Pharmaceutical Reaction System
  - Develop DRSMs & Discover Complex Stoichiometry
    - 10 Species involved in 8 reactions
    - Cannot Tell you Much at Present
- ExxonMobil Continuous Polymerization Process
  - Develop Meta-Models of KDM
    - Make Them More Accurate with Plant Data
    - USE for Optimization and Control Between SS Transition

# Dow Process: Batch Polymerization



- An Industrial Batch Reactor is Polymerizing
  - Fixed Amount of Monomer: $m_T$
  - In one Unit of Time: $t_b=1$
  - Product Quality Measures:
    - Impurities: $y_1(t_b)<1.00$
    - Unreacted Monomer: $y_1(t_b)<1.00$
  - Reactor Temperature: $0.14<T(t)<1.43$ (dimensionless)
  - Monomer Flow Rate: $0<u(t)<4.61$ (dimensionless)
  - Maximum Adiabatic Temperature: 1.714
  - Limit of cooling capability: 3.367

Values in Dimensionless Units for Proprietary reasons

# Optimization Objective: Productivity



- Design DoDE Runs: Vary Two (2) Inputs Over Time
  - Reactor Temperature, T(t) & Monomer Inflow, u(t)
- Process Constraints on T(t), u(t)
  - Adiabatic Temperature Constraint= ATC
  - Total Monomer is Constant
- Do NOT Violate Product Constraints
  - **$y_1(t_b)<1.00$ and $y_2(t_b)<1.00$**
- Optimization Objective:

Minimize batch Time ($t_b$) → Increase Productivity

# The "Chess" Game



- DoDE Side (CG) Knows LITTLE about Process:
  - Data from 3 Process runs: Data: u(t), T(t), $T_{ad}(t)$, $T_{c,in}$, $T_{c,out}$
- Dow Side: Does Experiments *in Silico* *
  - Reports Back: Product Qualities, Violation of Constraints
- Evolutionary Approach
  - Step 1: CG Designs Preliminary Set-1 (aim for 6.5%)
  - Step 2: Dow Performs Experiments
  - Step 3: CG Models, Optimizes, Designs Refined Runs
    - Optimize Process Further (aim for 20%)
  - Step 4: Dow Performs 2nd Set of Experiments
  - …



*DOW has a Knowledge-Driven Model for this Process

# First Cycle: DoDE Design A1



- Duration of the batch, ($t_b$): $t_b = 1 + 0.10A;\ -0.65 \le A \le 0$
- Temperature Profile (T(t)):

$$T(\tau_1) = 0.8\{1 + 0.07 w_1(\tau_1)\} = 0.8\{1 + 0.07[BP_0(\tau_1) + CP_1(\tau_1)]\}$$
$$= 0.8 + 0.056(B - C) + 0.112C\tau_1,\ \ \tau_1 = t / t_b$$
$$T(t) = 0.8 + 0.056(B - C) + 0.112C/(1 + 0.1A);\ 0 < t < 1 + 0.1A$$
$$-1 \le B \pm C \le 1 \Rightarrow -1 \le w_1(\tau_1) \le 1$$

- Monomer Feeding time ($t_m$): $t_m = 0.8 + 0.1F;\ -1 \le F \le 0$
- Monomer feeding (q(t)):

$$q(\tau_2) = 2.39\left(1 + 0.1[w_2(\tau_2)]\right) = 2.39\left(1 + 0.1[DP_0(\tau) + EP_1(\tau)]\right)$$
$$= 2.39\left(1 + 0.1(D - E) + 0.2E\tau_2\right);\ \ \tau_2 = t / t_m$$
$$q(t) = 2.39\left(1 + 0.1(D - E) + 0.2Et/(0.8 + 0.1F)\right);\ \ 0 < t, < 0.8 + 0.1F$$
$$-1 \le D \pm E \le 1 \Rightarrow -1 \le w_2(\tau_2) \le 1$$

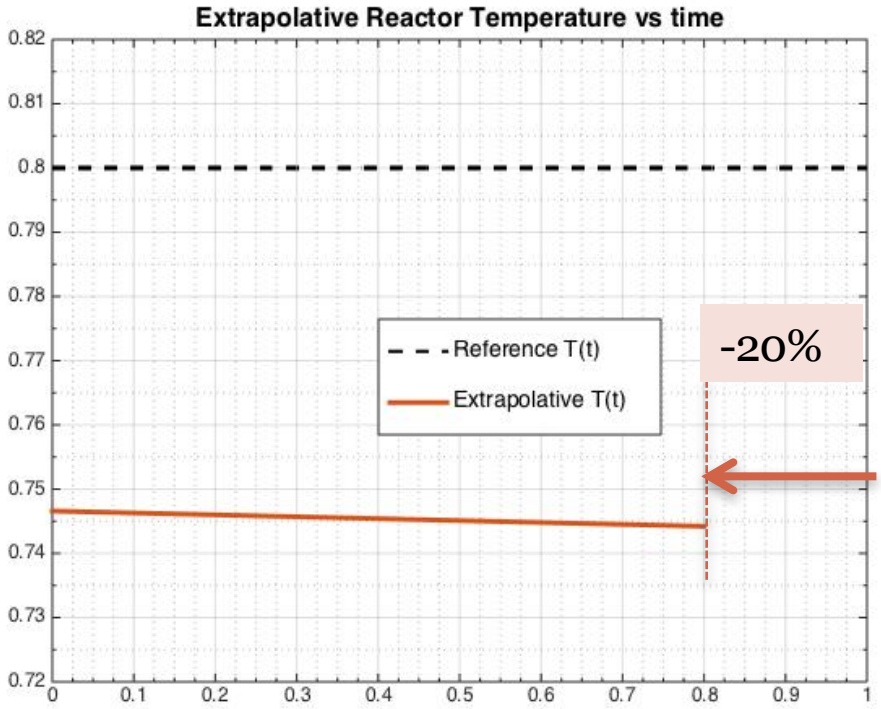
- Total Monomer Constraint: $F = -0.8D/(1 + 0.1D)$
- DoDE Design: D-Optimal Design for Linear RSM:

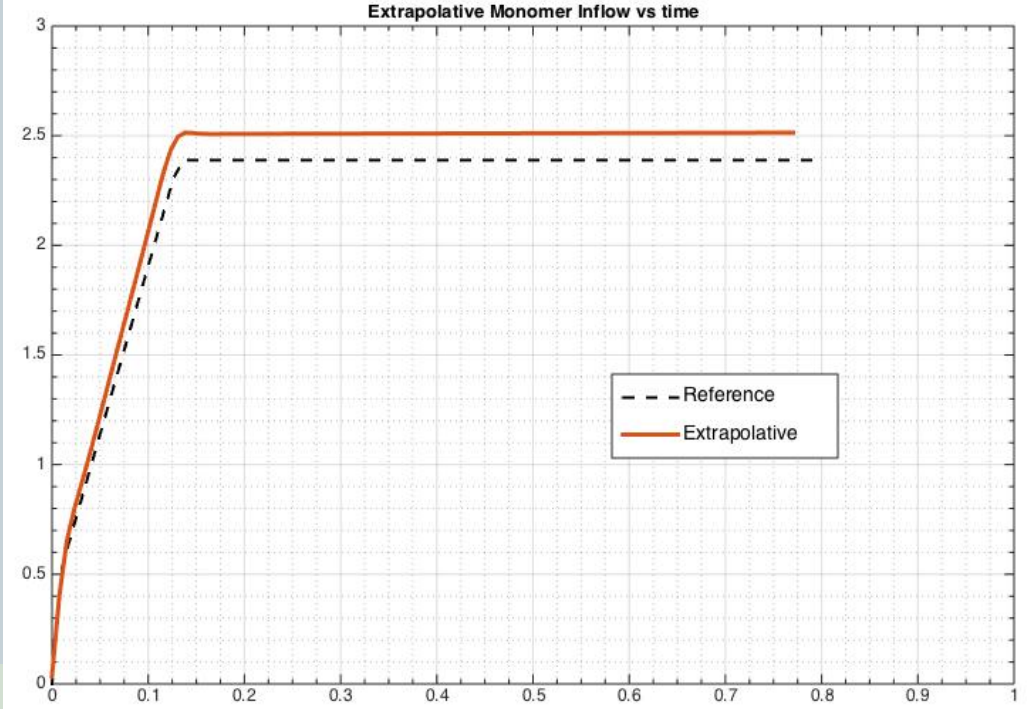
$$y_1 = \text{impurities} = \beta_{1,0} + \beta_{1,1}A + \beta_{1,2}B + \beta_{1,3}C + \beta_{1,4}D + \beta_{1,5}E$$
$$y_2 = \text{unreacted monomer} = \beta_{2,0} + \beta_{2,1}A + \beta_{2,2}B + \beta_{2,3}C + \beta_{2,4}D + \beta_{2,5}E$$

# Optimum Reached



- Decreasing $t_b$ from 1.0 to 0.8 (-20%) $-2 \le A \le 0$
  - Input Profile: *A=-2, B=-0.78, C=-0.017, D=0.51, E=0.016, F=-0.37*




*Outputs Expected Obtained* =
- Impurities: $y_1 = 0.94 \pm 0.03$ (0.87)
- Unreacted M: $y_2 = 0.32 \pm 0.05$ (0.13)

**Productivity Increased by** **20% !!!**

# What You Should Remember Tomorrow



- DoDE: First Generalization of DoE
  - Time-varying Inputs
- DRSM: Second Generalization of DoE
  - Using Time-Resolved Outputs
    - Excellent Approximation of Composition Profiles
  - Stepping Stone to a Kinetic Model
- Potential Benefits: ***Substantial***

## THANK YOU

MAY I ANSWER YOUR QUESTIONS?